

The Dot-Depth Hierarchy of Star-Free Languages is Infinite*

J. A. BRZOZOWSKI

*Department of Computer Science, University of Waterloo,
Waterloo, Ontario, Canada N2L 3G1*

AND

R. KNAST

Institute of Mathematics, Polish Academy of Sciences, 61-725 Poznań, Poland

Received May 10, 1977

Let A be a finite alphabet and A^* the free monoid generated by A . A language is any subset of A^* . Assume that all the languages of the form $\{a\}$, where a is either the empty word or a letter in A , are given. Close this basic family of languages under Boolean operations; let $\mathcal{B}^{(0)}$ be the resulting Boolean algebra of languages. Next, close $\mathcal{B}^{(0)}$ under concatenation and then close the resulting family under Boolean operations. Call this new Boolean algebra $\mathcal{B}^{(1)}$, etc. The sequence $\mathcal{B}^{(0)}, \mathcal{B}^{(1)}, \dots, \mathcal{B}^{(k)}, \dots$ of Boolean algebras is called the dot-depth hierarchy. The union of all these Boolean algebras is the family \mathcal{A} of star-free or aperiodic languages which is the same as the family of noncounting regular languages. Over an alphabet of one letter the hierarchy is finite; in fact, $\mathcal{B}^{(2)} = \mathcal{B}^{(1)}$. We show in this paper that the hierarchy is infinite for any alphabet with two or more letters.

INTRODUCTION

Let A be a finite, nonempty alphabet and A^* the free monoid generated by A , with identity 1 (the empty word). Elements of A^* are called words. The length of a word $x \in A^*$ is denoted by $|x|$. Note that $|1| = 0$. The concatenation of two words $x, y \in A^*$ is denoted by xy .

Any subset of A^* is called a language. If L_1 and L_2 are languages then $\bar{L}_1 = A^* - L_1$ is the complement of L_1 with respect to A^* , $L_1 \cup L_2$ is the union, and $L_1 \cap L_2$ is the intersection of L_1 and L_2 . Also $L_1 L_2 = \{w \in A^* \mid w = x_1 x_2, x_1 \in L_1, x_2 \in L_2\}$ is the concatenation or product of L_1 and L_2 .

For any family \mathcal{F} of languages let \mathcal{FM} be the smallest family of languages containing $\mathcal{F} \cup \{\{1\}\}$ and closed under concatenation. Similarly let \mathcal{FB} be the smallest family containing \mathcal{F} and closed under finite union and complementation. Thus \mathcal{FM} and \mathcal{FB} are the monoid and Boolean algebra, respectively, generated by \mathcal{F} .

* This work was supported in part by the National Research Council of Canada under Grant A-1617.

Let $\mathcal{L} = \{\{a\} \mid a \in A\}$; this is the finite family of languages whose elements are languages consisting of one word of length 1. We will write $\mathcal{L} \cup 1$ for $\mathcal{L} \cup \{\{1\}\}$. We use $\mathcal{L} \cup 1$ as the basic family of languages over the alphabet A . Now define the following sequence $\mathcal{B}^{(0)}, \mathcal{B}^{(1)}, \dots, \mathcal{B}^{(k)}, \dots$ of Boolean algebras:

$$\begin{aligned}\mathcal{B}^{(0)} &= (1 \cup \mathcal{L})B, \\ \mathcal{B}^{(k)} &= (\mathcal{B}^{(k-1)})MB = \mathcal{B}^{(0)}(MB)^k, \quad \text{for } k \geq 1.\end{aligned}$$

This sequence $(\mathcal{B}^{(0)}, \mathcal{B}^{(1)}, \dots, \mathcal{B}^{(k)}, \dots)$ is called the *dot-depth hierarchy*. A language L is of (dot) *depth* 0 iff $L \in \mathcal{B}^{(0)}$, and of *depth* k , $k \geq 1$, iff $L \in \mathcal{B}^{(k)} - \mathcal{B}^{(k-1)}$. Thus k is the minimum number of concatenation levels necessary to define L .

Let $\mathcal{A} = \bigcup_{k \geq 0} \mathcal{B}^{(k)}$; clearly \mathcal{A} is the smallest family containing $\mathcal{L} \cup 1$ and closed under Boolean operations and concatenation. This family is known as the aperiodic or star-free family [4, 5], and is identical to the family of noncounting regular languages [2, 4]. It was shown by Schützenberger [5] that $\mathcal{L} \subseteq A^*$ is star-free iff its syntactic monoid is finite and group-free, i.e., contains only one-element subgroups.

For languages over a one-letter alphabet one easily verifies that the dot-depth hierarchy is finite [1]. In fact, for $A = \{a\}$,

$$\mathcal{A}_a = (1 \cup \mathcal{L}_a)BMB = \mathcal{B}_a^{(1)},$$

where $\mathcal{L}_a = \{\{a\}\}$, \mathcal{A}_a is the family of aperiodic languages over a one-letter alphabet and $\mathcal{B}_a^{(1)}$ is the corresponding family of depth-one languages.

It was conjectured in [3] that the dot-depth hierarchy is infinite if the alphabet has two or more letters, i.e., that for each $k \geq 0$ there exists a language that is of depth $k+1$ but not of depth k . We prove this conjecture in this paper.

This paper is written by induction on k . In Sections 1–4 we treat the case $k = 1$ which provides the basis. The induction step consists of Sections 1⁺–4⁺.

I. BASIS: $k = 1$

1. DECOMPOSITIONS AND EQUIVALENCE RELATIONS

Let $(A^*)^n$ be the Cartesian product of n copies of A^* , for $n \geq 1$. Let $\pi_n: (A^*)^n \rightarrow A^*$ be defined as follows. For $X = (x_1, \dots, x_n) \in (A^*)^n$, $\pi_n(X) = x_1 \cdots x_n$. An *n-decomposition* is any element X of $(A^*)^n$. We say that X is an *n-decomposition* of $x \in A^*$ iff $\pi_n(X) = x$. Let $\Omega_n(x)$ be the set of all *n-decompositions* of x . Clearly $\Omega_n(x)$ is a finite set. For example, let $A = \{a, b\}$ and $x = aba$. Then x has the following 2-decompositions:

$$\Omega_2(x) = \{(1, aba), (a, ba), (ab, a), (aba, 1)\}.$$

DEFINITION 1. Let \sim be any equivalence relation on A^* . We define an equivalence relation \sim on $(A^*)^n$ derived from \sim on A^* as follows. If $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ then

$$X \sim Y \quad \text{iff} \quad x_i \sim y_i \quad \text{for } i = 1, \dots, n.$$

Let the equivalence class of \sim containing $x \in A^*$ be $[x]$. Similarly, let the class of \sim containing $X \in (A^*)^n$ be $[X]$. Clearly $[X] = [(x_1, \dots, x_n)]$ can be identified with $([x_1], \dots, [x_n])$. Let

$$\tilde{\Omega}_n(x) = \{[X] \mid X \in \Omega_n(x)\}$$

for all $x \in A^*$. Thus $\tilde{\Omega}_n(x)$ is "the set of all n -decompositions of x that are distinct with respect to the relation \sim ." For example, consider the equivalence defined by:

$$x \sim 1 \quad \text{iff} \quad x = 1,$$

and for $x \neq 1$,

$$x \sim y \quad \text{iff} \quad y \neq 1.$$

Under this equivalence $\tilde{\Omega}_2(aba) = \{([1], [a]), ([a], [a]), ([a], [1])\}$.

DEFINITION 2. Let \sim be any equivalence relation on A^* , $n \geq 1$ and $x, y \in A^*$.

(a) Define the binary relation \subset_n on A^* :

$$x \subset_n y \quad \text{iff} \quad \tilde{\Omega}_n(x) \subseteq \tilde{\Omega}_n(y).$$

(b) Define the equivalence relation \sim_n on A^* :

$$x \sim_n y \quad \text{iff} \quad x \subset_n y \text{ and } y \subset_n x.$$

We will say that an equivalence relation \sim on A^* is 1-pure iff $x \sim 1$ implies $x = 1$ for all $x \in A^*$.

PROPOSITION 1. For all $n \geq 1$ and $x, y, z_1, z_2 \in A^*$,

(a) \subset_n is reflexive and transitive.

(b) If \sim is 1-pure then

$$x \subset_n y \text{ implies } x \sim y \quad \text{and} \quad x \subset_{n+1} y \text{ implies } x \subset_n y.$$

(c) If \sim is a 1-pure congruence, then

$$x \subset_n y \text{ implies } z_1 x z_2 \subset_n z_1 y z_2.$$

Proof. (a) Obvious.

(b) Clearly $X = (x, 1, \dots, 1) \in \Omega_n(x)$. If $x \subset_n y$ there exists $Y \in \Omega_n(y)$, $Y = (y_1, \dots, y_n)$ such that $X \sim Y$. Since \sim is 1-pure, $Y = (y, 1, \dots, 1)$. Hence $x \sim y$.

To prove the second claim, suppose $X = (x_1, \dots, x_n) \in \Omega_n(x)$. Then $\hat{X} = (x_1, \dots, x_n, 1) \in \Omega_{n+1}(x)$. If $x \subset_{n+1} y$ and \sim is 1-pure, there exists $\hat{Y} = (y_1, \dots, y_n, 1) \in \Omega_{n+1}(y)$ such that $\hat{X} \sim \hat{Y}$. Then $Y = (y_1, \dots, y_n) \in \Omega_n(y)$ and $X \sim Y$. Therefore $x \subset_n y$.

(c) We will first show that $x \subset_n y$ implies $ax \subset_n ay$ for all $a \in A$. By induction on the length of z_1 it follows that $x \subset_n y$ implies $z_1 x \subset_n z_1 y$. The claim for z_2 follows by left-right symmetry.

Let $U = (u_1, \dots, u_n) \in \Omega_n(ax)$. Let u_i be the first component such that $|u_i| > 0$. Such a u_i always exists since $|ax| > 0$. The form of u_i must be $u_i = au$ for some $u \in A^*$. Thus $U = (1, \dots, 1, au, u_{i+1}, \dots, u_n)$. Let $X = (1, \dots, 1, u, u_{i+1}, \dots, u_n)$; clearly $X \in \Omega_n(x)$. By the hypothesis $x \subset_n y$ and 1-purity of \sim , there exists $Y = (1, \dots, 1, v, v_{i+1}, \dots, v_n) \in \Omega_n(y)$ such that $X \sim Y$. Note that $u \sim v$, and $au \sim av$ because \sim is a congruence. Let $V = (1, \dots, 1, av, v_{i+1}, \dots, v_n)$. Then $U \sim V$ and $V \in \Omega_n(ay)$. Therefore $ax \subset_n ay$. ■

PROPOSITION 2. For all $n \geq 1$ and $x, y \in A^*$,

- (a) If \sim is of finite index then so is \sim_n .
- (b) If \sim is 1-pure then so is \sim_n and

$$x \sim_{n+1} y \text{ implies } x \sim_n y.$$

- (c) If \sim is a 1-pure congruence then so is \sim_n .

Proof. (a) If \sim is of index i , then there are i^n n -decomposition classes. There are therefore $\leq 2^{i^n}$ sets of the form $\tilde{\Omega}_n(x)$.

(b) The fact that \sim_n is 1-pure is obvious, and the second claim follows directly from Proposition 1(b).

- (c) This follows directly from Proposition 1(c). ■

2. DECOMPOSITIONS AND CONCATENATION

From now on we assume that \sim is a 1-pure equivalence relation of finite index on A^* . Define

$$\mathcal{B}^{(0)} = \{L \subseteq A^* \mid L \text{ is a union of equivalence classes of } \sim\}.$$

Clearly $\mathcal{B}^{(0)}$ is a finite Boolean algebra with the equivalence classes $[x]$ as atoms. In this section we characterize $\mathcal{B}^{(0)}MB$ with the aid of \sim_n .

Denote by $[x]_n$ the equivalence class of \sim_n containing x . For $X \in \Omega_n(x)$ let

$$\pi_n[X] = [x_1] \cdots [x_n].$$

Here, each $[x_i]$ is viewed as a language and the multiplication is just concatenation of languages. Clearly

$$\pi_n[X] = \{z \in A^* \mid [X] \in \tilde{\Omega}_n(z)\}.$$

Define the languages $Y(x)$ and $N(x)$ (for *yes* and *no*):

$$Y(x) = \bigcap_{[X] \in \tilde{\Omega}_n(x)} \pi_n[X] \quad \text{and} \quad N(x) = \bigcap_{[X] \notin \tilde{\Omega}_n(x)} \overline{\pi_n[X]}.$$

PROPOSITION 3. $[x]_n = Y(x) \cap N(x)$.

Proof. If $z \in [x]_n$ then $\tilde{\Omega}_n(z) = \tilde{\Omega}_n(x)$. Thus $[X] \in \tilde{\Omega}_n(x)$ implies $[X] \in \tilde{\Omega}_n(z)$ and $z \in \pi_n[X]$. Therefore $z \in Y(x)$. Similarly if $[X] \notin \tilde{\Omega}_n(x)$ then $z \notin \pi_n[X]$ and $z \in \overline{\pi_n[X]}$. Therefore $z \in N(x)$.

Conversely $z \in Y(x) \cap N(x)$ implies $z \in \pi_n[X]$ iff $[X] \in \tilde{\Omega}_n(x)$. Hence $\tilde{\Omega}_n(z) = \tilde{\Omega}_n(x)$ and $z \in [x]_n$. ■

Corresponding to each n define the family:

$$\mathcal{B}_n = \{L \subseteq A^* \mid L \text{ is a union of equivalence classes of } \sim_n\}.$$

Again \mathcal{B}_n is a finite Boolean algebra, \sim_n being of finite index. Let

$$\mathcal{B}^{(1)} = \bigcup_{n \geq 1} \mathcal{B}_n.$$

PROPOSITION 4. For all $n \geq 1$,

- (a) $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$.
- (b) $\mathcal{B}_n = (\mathcal{B}^{(0)})^n B$. Hence $\mathcal{B}^{(0)} \subseteq \mathcal{B}_n$.
- (c) $\mathcal{B}^{(1)} = \mathcal{B}^{(0)} MB$, i.e., $\bigcup_{n \geq 1} \mathcal{B}_n = \bigcup_{n \geq 1} ((\mathcal{B}^{(0)})^n B) = (\bigcup_{n \geq 1} (\mathcal{B}^{(0)})^n) B$.

Proof. (a) This follows directly from Proposition 2(b).

(b) Suppose $L \in \mathcal{B}_n$. Since $(\mathcal{B}^{(0)})^n B$ is a Boolean algebra, it suffices to show that each $[x]_n$ is in $(\mathcal{B}^{(0)})^n B$. By Proposition 3, $[x]_n$ is a Boolean function of elements $\pi_n[X]$ from $(\mathcal{B}^{(0)})^n$. Hence $\mathcal{B}_n \subseteq (\mathcal{B}^{(0)})^n B$.

Conversely it is enough to show that $L \in (\mathcal{B}^{(0)})^n$ implies $L \in \mathcal{B}_n$, since \mathcal{B}_n is a Boolean algebra. In fact, any $L \in (\mathcal{B}^{(0)})^n$ can be expressed as a finite union of languages of the form $[x_1] \cdots [x_n] = \pi_n[X]$, since concatenation distributes over union. Thus we need to show only that $\pi_n[X] \in \mathcal{B}_n$ for all $X \in \Omega_n(x)$. We claim that

$$\pi_n[X] = \bigcup_{w \in J} [w]_n, \quad (1)$$

where $J = \{z \mid [X] \in \tilde{\Omega}_n(z)\}$. For suppose $y \in \pi_n[X]$. Then $y = y_1 \cdots y_n$, $y_i \in [x_i]$, $i = 1, \dots, n$. Let $Y = (y_1, \dots, y_n)$; then $[X] = [Y]$. Thus $y \in \pi_n[X]$ implies $[X] \in \tilde{\Omega}_n(y)$, i.e., $y \in J$. But then $y \in \bigcup_{w \in J} [w]_n$.

On the other hand, suppose $y \in [w]_n$ for some $w \in J$. Now $[w]_n = Y(w) \cap N(w)$ and $\pi_n[X]$ appears in $Y(w)$ since $[X] \in \tilde{\Omega}_n(w)$. Thus $y \in [w]_n$ implies $y \in Y(w)$ and $y \in \pi_n[X]$. This completes the proof of the claim (1). By (1), $\pi_n[X] \in \mathcal{B}_n$ and $(\mathcal{B}^{(0)})^n \subseteq \mathcal{B}_n$.

(c) $L \in \mathcal{B}^{(1)}$ implies $L \in \mathcal{B}_n$ for some n and by (b) $\mathcal{B}_n = (\mathcal{B}^{(0)})^n B \subseteq \mathcal{B}^{(0)} MB$. Thus $\mathcal{B}^{(1)} \subseteq \mathcal{B}^{(0)} MB$. Conversely $L \in \mathcal{B}^{(0)} MB$ implies $L \in (\mathcal{B}^{(0)})^n B$ for some n and $(\mathcal{B}^{(0)})^n B = \mathcal{B}_n$. Thus $L \in \mathcal{B}^{(0)} MB$ implies $L \in \mathcal{B}_n \subseteq \mathcal{B}^{(1)}$. Hence $\mathcal{B}^{(0)} MB \subseteq \mathcal{B}^{(1)}$. ■

In summary, if a family $\mathcal{B}^{(0)}$ of languages is defined by an equivalence relation \sim , then the family $(\mathcal{B}^{(0)})^n B$ is defined by \sim_n .

3. LANGUAGES OF DOT-DEPTH 1

Let \sim be the largest 1-pure equivalence on A^* for any A . Then there are only two equivalence classes $[1] = \{1\}$ and $[a] = A^+$, $a \in A$. Now let $\mathcal{B}^{(0)}$ be the family defined by \sim , i.e.,

$$\mathcal{B}^{(0)} = \{\phi, \{1\}, A^+, A^*\}.$$

One verifies that the equivalence classes of \sim_n are:

$$\begin{aligned} [1]_n &= 1, \\ [a]_n &= A, \\ [a^2]_n &= A^2, \\ &\dots \\ [a^{n-1}]_n &= A^{n-1}, \\ [a^n]_n &= A^n A^*. \end{aligned}$$

Now it is easily seen that $\mathcal{B}^{(1)} = \mathcal{B}^{(0)}MB = \bigcup_{n \geq 1} (\mathcal{B}^{(0)})^n B$ is closed under concatenation. Thus $\mathcal{B}^{(2)} = \mathcal{B}^{(1)}$. In the case of a one-letter alphabet $A = \{a\}$, this means that $\mathcal{A} = \mathcal{B}^{(1)}$, i.e., *a language over a one-letter alphabet is star-free iff it is of depth 0 or 1*.

We now consider the case of two or more letters.

From now on \sim represents the following equivalence:

- (a) If $x \in 1 \cup A$ then $x \sim y$ iff $x = y$.
- (b) If $x \notin 1 \cup A$ then $x \sim y$ iff $y \notin 1 \cup A$.

This is the largest equivalence relation on A^* that is pure for all $a \in 1 \cup A$ in the sense that $a \sim x$ implies $a = x$ for all $a \in 1 \cup A$. If the cardinality of A is $\#A$, the index of \sim is $\#A + 2$. One easily verifies that \sim is a congruence. We will call this the *2-pure congruence* meaning that $x \sim y$ implies $x = y$ for $|x| < 2$.

LEMMA 1. For all $n \geq 1$, $y \in A^*$,

$$y^{2n} \sim_n y^{2n+1}.$$

Proof. We first show that $\tilde{\Omega}_n(y^{2n+1}) \subseteq \tilde{\Omega}_n(y^{2n})$. There is nothing to prove if $y = 1$. Now suppose $y = a$, where $a \in A$. Let $U = (u_1, \dots, u_n) \in \Omega_n(y^{2n+1})$. There must be at least one $u_i = a^s$ with $s \geq 3$. Otherwise

$$|y^{2n+1}| = |a^{2n+1}| = 2n + 1 = \sum_{i=1}^n |u_i| \leq 2n,$$

a contradiction. Let $u_i' = a^{s-1}$. Since $|a^{s-1}| \geq 2$, $a^s \sim a^{s-1}$. Let $U' = (u_1, \dots, u_{i-1}, u_i', u_{i+1}, \dots, u_n)$. Then $\pi_n(U') = a^{2n}$ and $U' \sim U$. Thus $a^{2n+1} \subset_n a^{2n}$.

Assume now that $|y| \geq 2$. First suppose that $|u_i| \geq |y|$ for all i . Then all u_i in U must be of the form $u_i = y_1 y^s y_2$ where y_2 is a prefix of y , y_1 is a suffix of y , and $s \geq 0$. If there exists a u_i with $s \geq 2$, then $|y_1 y^s y_2| \geq 2$ and $|y_1 y^{s-1} y_2| \geq 2$, i.e., $y_1 y^s y_2 \sim y_1 y^{s-1} y_2$. If there exists a u_i with $s = 1$ and $|y_1 y_2| \geq 2$ again $y_1 y^s y_2 \sim y_1 y^{s-1} y_2$. Therefore, assume that for all u_i either $s = 1$ and $|y_1 y_2| \leq 1$ or $s = 0$. In the first

case $|u_i| = |y_1 y y_2| \leq |y| + 1$. In the second case $|y_1 y_2| \leq 2|y|$. In both cases $|u_i| \leq 2|y|$. Hence $|y^{2n+1}| = (2n+1)|y| = \sum_{i=1}^n |u_i| \leq 2n|y|$, a contradiction. Finally, if there exists a u_j with $|u_j| < |y|$, then there also exists a u_k with $|u_k| > 2|y|$. This u_k must be of the form $u_k = y_1 y^s y_2$, where either $s > 1$ or $s = 1$ and $|y_1 y_2| > |y| \geq 2$, and we proceed as above. Therefore, one can always find $U' \in \Omega_n(y^{2n})$ such that $U' \sim U$. We have therefore shown that $y^{2n+1} C_n y^{2n}$.

The argument for $y^{2n} C_n y^{2n+1}$ is essentially the same except we insert y instead of removing it. For $y = a$, there must be a u_i with $|u_i| \geq 2$. Then $u_i = a^s$, $s \geq 2$ and $a^s \sim a^{s+1}$. For $|y| \geq 2$, there must exist $u_i = y_1 y^s y_2$ with $|u_i| \geq 2$. Then $y_1 y^s y_2 \sim y_1 y^{s+1} y_2$. ■

LEMMA 2. Let \sim be the 2-pure congruence on A^* , let $n \geq 1$ and $x, y \in A^*$. Then

$$|x| > n \text{ implies } x C_n xyx.$$

Proof. Let $X = (x_1, \dots, x_n) \in \Omega_n(x)$. Let x_i be such that $|x_i| \geq 2$; such an x_i always exists since $|x| = \sum_{i=1}^n |x_i| > n$. Let $Y = (x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n)$ where $x_i' = x_i \cdots x_n y x_1 \cdots x_i$. Then $|x_i'| \geq 2$, $x_i \sim x_i'$ and $X \sim Y$. Since $\pi_n(Y) = xyx$, we have $x C_n xyx$. ■

LEMMA 3. Let $x, y, z \in A^*$, $n \geq 1$, and $|x| > n$. Then

$$x(yxzx)^{2n} \sim_n x(zxyx)^{2n}.$$

Proof. Let $u = x(yxzx)^{2n}$. By Lemma 1,

$$u \sim_n u' = x(yxzx)^{2n+1} = yxzx(yxzx)^{2n-1} yxzx.$$

Let $w = zx(yxzx)^{2n-1}y$. Then $u \sim_n (xyx)w(xzx)$. Let $v = x(zxyx)^{2n} = xzx(yxzx)^{2n-1}yx = xzxw$. By Lemma 2, $x C_n xyx$ and $x C_n xzx$. By transitivity of C_n , $v = xzxw C_n xyxw C_n xyxwzx = u' \sim_n u$. Thus $v C_n u$ and, by symmetry, $u C_n v$. Therefore $u \sim_n v$. ■

We now give an example of a language that is not in $\mathcal{B}^{(1)}$. Let $A_2 = \langle A, Q, q_1, F, \tau \rangle$ be the finite automaton of Fig. 1, where $A = \{a, b\}$ is the alphabet, $Q = \{0, 1, 2, 3\}$

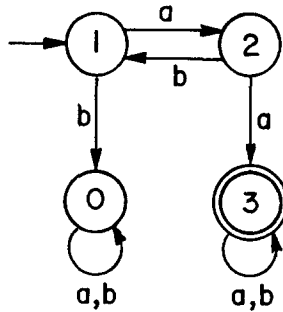


FIG. 1. Automaton A_2 .

is the set of states, $q_1 = 1$ is the initial state, $F = \{3\}$ is the set of final states, and τ is the transition function given by Fig. 1. One verifies that \mathbf{A}_2 is reduced. Let L_2 be the language recognized by \mathbf{A}_2 , $L_2 = (ab)^* aaA^*$.

PROPOSITION 5. $L_2 \in \mathcal{B}^{(2)} - \mathcal{B}^{(1)}$, i.e., L_2 is a depth-2 language.

Proof. Suppose $L_2 \in \mathcal{B}^{(1)}$. Then L_2 is a union of congruence classes of \sim_n for some $n \geq 1$. Let $x = (ab)^n$, $y = a$ and $z = b$. One easily verifies that

$$x(yxzx)^{2n} \in L_2 \quad \text{and} \quad x(zxyx)^{2n} \notin L_2.$$

But by Lemma 3, $x(yxzx)^{2n} \sim_n x(zxyx)^{2n}$, and these two words are in the same congruence class. This is a contradiction. Hence $L_2 \notin \mathcal{B}^{(1)}$.

In automaton \mathbf{A}_2 , let $Z_i = \{w \in A^* \mid \tau(1, w) = i\}$, and let $D_1 = (ab)^*$. Then, from Fig. 1,

$$\begin{aligned} Z_0 &= D_1 b A^*, \\ Z_1 &= D_1, \\ Z_2 &= D_1 a, \\ L_2 &= Z_3 = (D_1 a) a A^*, \end{aligned}$$

and $\bar{D}_1 = bA^* \cup A^*bbA^* \cup A^*a \cup A^*aaA^*$, showing that $D_1 \in \mathcal{B}^{(1)}$, since $A^* = \bar{\phi}$ is in $\mathcal{B}^{(0)}$.

It now follows that $L_2 = D_1 a^2 A^*$ is in $\mathcal{B}^{(2)}$. Altogether L_2 is a language of depth 2. ■

4. ON SYNTACTIC SEMIGROUPS OF DEPTH-ONE LANGUAGES

Let $L \subseteq A^+$ be a language. The syntactic congruence of L is defined as follows. For $x, y \in A^+$,

$$x \equiv_L y \quad \text{iff for all } u, v \in A^*, \quad uxv \in L \Leftrightarrow uyv \in L.$$

Let $S_L = A^+ / \equiv_L$ be the quotient semigroup of A^+ modulo the congruence \equiv_L ; S_L is called the syntactic semigroup of L [4]. Let $\mu: A^+ \rightarrow S_L$ be the natural morphism associating with each $x \in A^+$, the equivalence class of \equiv_L containing x . We will denote by \bar{x} the image of x under μ (i.e., $\mu(x) = \bar{x}$).

We will say that a semigroup S is *aperiodic* iff there exists $m \geq 1$ such that $f^m = f^{m+1}$ for all $f \in S$. We say that S is *1-mutative* iff there exists $m \geq 1$ such that

$$(fg)^m = (gf)^m,$$

for all $f, g \in S$. The two conditions are equivalent to S being \mathcal{J} -trivial if S is finite [6]. The reasons for our choice of terminology will become clearer in the induction step.

The following gives a necessary condition for membership in $\mathcal{B}^{(1)}$.

PROPOSITION 6. Let $L \subseteq A^+$ and let S_L be the syntactic semigroup of L .

- (a) If $L \in \mathcal{B}^{(1)}$ then for each idempotent $e \in S_L$, $eS_L e$ is finite, aperiodic, and 1-mutative.
- (b) Suppose S_L is a monoid. Then $L \in \mathcal{B}^{(1)}$ implies that S_L is finite, aperiodic, and 1-mutative.

Proof. (a) If $L \in \mathcal{B}^{(1)}$, then L is a union of congruence classes of \sim_n for some $n \geq 1$. Since \sim_n is of finite index, S_L is finite. Since S_L is the image of A^+ under μ , there exists $y \in A^+$ such that $y = f$ for each $f \in S_L$. By Lemma 1

$$y^{2n} \sim_n y^{2n+1}. \quad (2)$$

Since L is a union of congruence classes of \sim_n it follows that $x \sim_n x'$ implies $\underline{x} = \underline{x'}$ for all $x, x' \in A^+$. Therefore by (2)

$$f^{2n} = f^{2n+1}. \quad (3)$$

(The reader should note that we have just shown that if L is in $\mathcal{B}^{(1)}$ then its syntactic semigroup S_L satisfies (3) for all $f \in S_L$, i.e., is group-free [4].)

Now let $e, f, g \in S_L$, let e be an idempotent, and let $u, x, y, z \in A^+$ be such that $\underline{u} = e$, $y = f$, $z = g$, and $x = u^{n+1}$. By Lemma 3,

$$x(yxzx)^{2n} \sim_n x(zxyx)^{2n}, \quad (4)$$

and

$$e(fege)^{2n} = e(gefe)^{2n}. \quad (5)$$

From (3) and (5) it follows that $eS_L e$ satisfies the required conditions with $m = 2n$, since

$$((efe)(ege))^m = e(fege)^m = e(gefe)^m = ((ege)(efe))^m. \quad (6)$$

(b) Let 1 be the identity of S_L . Since (6) holds for all idempotents, it holds for $e = 1$ and we have $(fg)^m = (gf)^m$. This and (3) show that S_L is 1-mutative and aperiodic.

These results were obtained first by Simon [6] by different means. He also showed the converse of (b), i.e.:

(b') Suppose S_L is a monoid. If S_L is finite, aperiodic, and 1-mutative then $L \in \mathcal{B}^{(1)}$.

This concludes the basis.

II. INDUCTION STEP: $k > 1$

1+. DECOMPOSITIONS AND GENERALIZED EQUIVALENCE RELATIONS

We now assume that Section 1 corresponds to $k = 1$, and we generalize all the notions by induction on k . The induction hypothesis is that everything has been done for k , and we consider $k + 1$.

DEFINITION 1+. For each $k \geq 1$, $n \geq 1$ let \sim_n^k be an equivalence relation on A^* .

We define a relation \sim^{k+1} on $(A^*)^n$ derived from \sim_n^k as follows. If $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ then

$$\begin{aligned} k = 0: \quad X &\sim^1 Y && \text{iff } X \sim Y \text{ as in Definition 1,} \\ k > 0: \quad X &\sim^{k+1} Y && \text{iff } x_i \sim_n^k y_i \text{ for } i = 1, \dots, n. \end{aligned}$$

Let the equivalence class of \sim_n^k containing $x \in A^*$ be $[x]_n^k$. Similarly let the class of \sim^k containing $X = (x_1, \dots, x_n) \in (A^*)^n$ be $[X]^k$. Clearly $[X]^{k+1}$ can be identified with $([x_1]_n^k, \dots, [x_n]_n^k)$. Let

$$\tilde{\Omega}_n^k(x) = \{[X]^k \mid X \in \Omega_n(x)\},$$

for all $x \in A^*$.

DEFINITION 2⁺. Let \sim be any equivalence relation on A^* , $n, k \geq 1$ and $x, y \in A^*$.

(a) Define a binary relation \subset_n^k on A^* :

$$\begin{aligned} k = 1: \quad \subset_n^1 &= \subset_n \text{ of Definition 2,} \\ k > 1: \quad x \subset_n^k y &\quad \text{iff } \tilde{\Omega}_n^k(x) \subseteq \tilde{\Omega}_n^k(y). \end{aligned}$$

(b) Define the equivalence relation \sim_n^k on A^* :

$$\begin{aligned} k = 1: \quad \sim_n^1 &= \sim \text{ of Definition 2,} \\ k > 1: \quad x \sim_n^k y &\quad \text{iff } x \subset_n^k y \text{ and } y \subset_n^k x. \end{aligned}$$

To illustrate this inductive procedure, we have the following order in which the concepts appear:

- (1) $x \sim_n^1 y$ is defined in the basis.
- (2) $X \sim^2 Y$ iff $x_i \sim_n^1 y_i$ for all $i = 1, \dots, n$ (Definition 1⁺).
- (3) This yields $[X]^2$ and $\tilde{\Omega}_n^2(x)$.
- (4) $x \subset_n^2 y$ iff $\tilde{\Omega}_n^2(x) \subseteq \tilde{\Omega}_n^2(y)$.
- (5) $x \sim_n^2 y$ iff $x \subset_n^2 y$ and $y \subset_n^2 x$.

Thus we have gone through the full cycle.

PROPOSITION 1⁺. Let $n, k \geq 1$ and $x, y, z_1, z_2 \in A^*$.

- (a) \subset_n^k is reflexive and transitive.
- (b) If \sim is 1-pure then

$$x \subset_n^{k+1} y \text{ implies } x \sim_n^k y \quad \text{and} \quad x \subset_{n+1}^k y \text{ implies } x \subset_n^k y.$$

(c) If \sim is a 1-pure congruence, then

$$x \overset{k}{\subset}_n y \text{ implies } z_1 x z_2 \overset{k}{\subset}_n z_1 y z_2.$$

Proof. (a) Trivial.

(b) $k = 1$: Proposition 1(b).

$k > 1$: Clearly $X = (x, 1, \dots, 1) \in \Omega_n(x)$. If $x \overset{k}{\subset}_{n+1} y$ there exists $Y = (y_1, \dots, y_n) \in \Omega_n(y)$ such that $X \sim^{k+1} Y$. Since \sim_n^k is 1-pure by the inductive assumption (Proposition 2+), Y is of the form $Y = (y, 1, \dots, 1)$ and $x \sim_n^k y$.

For the second claim, suppose $X = (x_1, \dots, x_n) \in \Omega_n(x)$. Then $\hat{X} = (x_1, \dots, x_n, 1) \in \Omega_{n+1}(x)$. If $x \overset{k}{\subset}_{n+1} y$ and \sim is 1-pure there exists $\hat{Y} = (y_1, \dots, y_n, 1)$ such that $\hat{X} \sim^k \hat{Y}$ and $\hat{Y} \in \Omega_{n+1}(y)$. Then $Y = (y_1, \dots, y_n) \in \Omega_n(y)$ and $X \sim^k Y$. Therefore $x \overset{k}{\subset}_n y$.

(c) Same argument as in Proposition 1(c). ■

PROPOSITION 2+. For all $n, k \geq 1$ and $x, y \in A^*$:

(a) If \sim is of finite index then so is \sim_n^k .

(b) If \sim is 1-pure, then so is \sim_n^k and

$$x \overset{k}{\sim}_{n+1} y \text{ implies } x \overset{k}{\sim}_n y.$$

(c) If \sim is a 1-pure congruence then so is \sim_n^k .

Proof. Same as Proposition 2 after \sim_n is replaced by \sim_n^k . ■

2+. DECOMPOSITIONS AND REPEATED CONCATENATION

Again \sim is assumed to be a 1-pure equivalence relation of finite index. Denote by $[x]_n^k$ the class of \sim_n^k containing x , and for $X \in \Omega_n(x)$ let

$$\pi_n[X]^{k+1} = [x_1]_n^k \cdots [x_n]_n^k.$$

We have

$$\pi_n[X]^{k+1} = \{z \in A^* \mid [X]^{k+1} \in \bar{\Omega}_n^{k+1}(z)\}.$$

Define also

$$Y^k(x) = \bigcap_{[X]^k \in \bar{\Omega}_n^k(x)} \pi_n[X]^k \quad \text{and} \quad N^k(x) = \bigcap_{[X]^k \notin \bar{\Omega}_n^k(x)} \pi_n[X]^k.$$

PROPOSITION 3+. $[x]_n^k = Y^k(x) \cap N^k(x)$.

Proof. Repeat the proof of Proposition 3 with \sim_n^k instead of \sim_n . ■

Corresponding to each \sim_n^k define:

$$\mathcal{B}_n^{(k)} = \{L \subseteq A^* \mid L \text{ is a union of equivalence classes of } \sim_n^k\}.$$

Again $\mathcal{B}_n^{(k)}$ is a finite Boolean algebra. Let

$$\mathcal{B}^{(k)} = \bigcup_{n \geq 1} \mathcal{B}_n^{(k)},$$

PROPOSITION 4⁺. For all $n, k \geq 1$,

- (a) $\mathcal{B}_n^{(k)} \subseteq \mathcal{B}_{n+1}^{(k)}$,
- (b) $\mathcal{B}_n^{(k+1)} = (\mathcal{B}_n^{(k)})^n B$, hence $\mathcal{B}_n^{(k)} \subseteq \mathcal{B}_n^{(k+1)}$,
- (c) $\mathcal{B}^{(k+1)} = (\mathcal{B}^{(k)}) MB = \mathcal{B}^{(0)}(MB)^{k+1}$.

Proof. Repeat the proof of Proposition 4 with \sim_n^k instead of \sim_n . ■

It follows that the family of aperiodic languages is

$$\mathcal{A} = \bigcup_{k \geq 0} \mathcal{B}^{(k)}.$$

3⁺. LANGUAGES OF DOT-DEPTH k

Again, let \sim be the 2-pure congruence.

LEMMA 1⁺. For all $n, k \geq 1$, $y \in A^*$, there exists $m \geq 1$ such that $y^m \sim_n^k y^{m+1}$.

Proof. Let $m_k = 2n(\sum_{i=0}^{k-1} n^i)$ for $k \geq 1$. We claim that $y^{m_k} \sim_n^k y^{m_k+1}$.

$k = 1$: We have $m_1 = 2n$ and the result holds by Lemma 1.

$k > 1$: Assume the result holds for k , and that $|y| \geq 1$.

Let $U = (u_1, \dots, u_n) \in \Omega_n(y^{m_{k+1}+1})$. Then there exists at least one u_i such that

$$\begin{aligned} |u_i| &> \frac{m_{k+1}}{n} |y| = 2 \left(\sum_{i=0}^{k-1} n^i \right) |y| = \left(2n \left(\sum_{i=0}^{k-1} n^i \right) + 2 \right) |y| \\ &= (m_k + 2) |y|. \end{aligned}$$

Now u_i must be of the form $u_i = y_1 y^s y_2$ where $|y_1 y_2| \leq 2|y|$. Hence $s > m_k$ and by the induction hypothesis $y^s \sim_n^k y^{s-1}$. Let $U' = (u_1, \dots, u_{i-1}, u_i', u_{i+1}, \dots, u_n)$ where $u_i' = y_1 y^{s-1} y_2$. Then $u_i \sim_n^k u_i'$ and $U \sim^{k+1} U'$. Since $\pi_n(U') = y^{m_{k+1}}$, we have $y^{m_{k+1}+1} \subseteq_n^{k+1} y^{m_{k+1}}$.

To prove $y^{m_{k+1}} \subseteq_n^{k+1} y^{m_{k+1}+1}$, use a similar argument, replacing y^s by y^{s+1} instead of y^{s-1} . ■

LEMMA 2⁺. Let $k \geq 0$, $n \geq 1$, $x, y \in A^*$, $|x| > n$. Define

$$u_0 = x$$

and

$$u_k = u_{k-1}(yu_{k-1}zu_{k-1})^{m_{k+1}}, \quad \text{for } k > 0,$$

where m_k is defined in Lemma 1⁺. Then

$$u_k \overset{k+1}{\underset{n}{\subset}} u_k y u_k \quad \text{and} \quad u_k \overset{k+1}{\underset{n}{\subset}} u_k z u_k.$$

Proof. $k = 0$: This reduces to Lemma 2.

$k > 0$: Let $w = yu_{k-1}zu_{k-1}$. We must show

$$u_k = u_{k-1}w^{m_{k+1}} \overset{k+1}{\underset{n}{\subset}} u_{k-1}w^{m_{k+1}}yu_{k-1}w^{m_{k+1}}. \quad (7)$$

Because of Proposition 1⁺(c) it is enough to show that

$$w^{m_{k+1}} \overset{k+1}{\underset{n}{\subset}} w^{m_{k+1}}yu_{k-1}w^{m_{k+1}} = v. \quad (8)$$

Let $W = (w_1, \dots, w_n) \in \Omega_n(w^{m_{k+1}})$. There must exist w_i such that $|w_i| \geq (m_{k+1}/n)|w| = (m_k + 2)|w|$. Also w_i must be of the form $w'w^s w''$, where w' is a suffix and w'' is a prefix of w . It follows that $s \geq m_k$. Hence

$$w^s \overset{k}{\underset{n}{\sim}} w^{m_k} \overset{k}{\underset{n}{\sim}} w^{2m_k+1} = w^{m_k}yu_{k-1}zu_{k-1}w^{m_k} = p.$$

Now we have the inductive assumption:

$$u_{k-1} \overset{k}{\underset{n}{\subset}} u_{k-1}yu_{k-1} \quad \text{and} \quad u_{k-1} \overset{k}{\underset{n}{\subset}} u_{k-1}zu_{k-1}.$$

Therefore

$$q = w^{m_k}yu_{k-1}w^{m_k} \overset{k}{\underset{n}{\subset}} w^{m_k}(u_{k-1}zu_{k-1})w^{m_k} = p.$$

On the other hand,

$$q \overset{k}{\underset{n}{\sim}} w^{m_{k+1}}yu_{k-1}w^{m_k} = w^{m_k}(yu_{k-1}zu_{k-1})yu_{k-1}w^{m_k}$$

and

$$p = w^{m_k}yu_{k-1}zu_{k-1}w^{m_k} \overset{k}{\underset{n}{\subset}} w^{m_k}yu_{k-1}z(u_{k-1}yu_{k-1})w^{m_k} \overset{k}{\underset{n}{\sim}} q.$$

Thus $p \overset{k}{\underset{n}{\sim}} q$, showing that

$$w^s \overset{k}{\underset{n}{\sim}} w^{m_k}yu_{k-1}w^{m_k} = q.$$

By Lemma 1⁺,

$$w^s \overset{k}{\underset{n}{\sim}} w^{m_{k+1}}yu_{k-1}w^{m_{k+1}}.$$

Now let $w'_i = w'w^{m_{k+1}}yu_{k-1}w^{m_{k+1}}w''$, and let $W' = (w_1, \dots, w_{i-1}, w'_i, w_{i+1}, \dots, w_n)$. Then $\pi_n(W')$ is of the form $w'w^{m_{k+1}}yu_{k-1}w^{m_{k+1}}w''$ which is \sim_n^{k+1} equivalent to $w^{m_{k+1}}yu_{k-1}w^{m_{k+1}} = v$. Now $W' \sim^{k+1} W$; i.e., we have shown that $w^{m_{k+1}} \subset_n^{k+1} v$. This is (8), and (7) follows.

To prove $u_k \subset_n^{k+1} u_k zu_k$ use a very similar argument, except that we show that

$$w^{m_k} \sim_n^k w^{m_k} zu_{k-1} w^{m_k} = v.$$

This holds since

$$w^{m_k} \sim_n^k w^{m_k} y u_{k-1} z u_{k-1} w^{m_k} \subset_n^k w^{m_k} y u_{k-1} z (u_{k-1} z u_{k-1}) w^{m_k} \sim_n^k v,$$

and

$$v \sim_n^k w^{m_k} y u_{k-1} z (u_{k-1}) z u_{k-1} w^{m_k} \subset_n^k w^{m_k} y u_{k-1} z (u_{k-1} y u_{k-1}) z u_{k-1} w^{m_k} \sim_n^k w^{m_k}. \quad \blacksquare$$

LEMMA 3⁺. Let $n, k \geq 1$, $|x| > n$, and $x, y, z \in A^*$. Let $u_0 = x$ and for $k \geq 1$, let

$$u_k = u_{k-1}(yu_{k-1}zu_{k-1})^m \quad \text{and} \quad v_k = u_{k-1}(zu_{k-1}yu_{k-1})^m.$$

Then m can be chosen in such a way that $u_k \sim_n^k v_k$.

Proof. $k = 1$: This is Lemma 3.

$k > 1$: Let $m = m_{k+1}$; then Lemmas 1⁺ and 2⁺ hold for \sim_n^{k+1} and \subset_n^{k+1} , respectively. By Lemma 1⁺ $u_{k+1} \sim_n^{k+1} u_k(yu_k zu_k)^{m+1} = u_k y u_k z u_k (yu_k zu_k)^{m-1} y u_k z u_k$. Let $w_k = zu_k(yu_k zu_k)^{m-1} y$. Then $u_{k+1} \sim_n^{k+1} (u_k y u_k) w_k (u_k z u_k)$. Also, $v_{k+1} = u_k w_k u_k$. By Lemma 2⁺, $u_k \subset_n^{k+1} u_k y u_k$ and $u_k \subset_n^{k+1} u_k z u_k$. Hence $u_{k+1} \subset_n^{k+1} v_{k+1}$. Similarly, $v_{k+1} \subset_n^{k+1} u_{k+1}$ and the result follows. \blacksquare

We now give an example for each $k \geq 1$ of a language that is not in $\mathcal{B}^{(k)}$. Let $\mathbf{A}_{k+1} = \langle A, Q, q_1, F, \tau \rangle$, where $A = \{a, b\}$, $Q = \{0, 1, \dots, k+2\}$, $q_1 = 1$, $F = \{k+2\}$ and for $i = 1, \dots, k+1$

$$\begin{aligned} \tau(i, a) &= i+1, & \tau(i, b) &= i-1, \\ \tau(0, a) &= \tau(0, b) = 0, \\ \tau(k+2, a) &= \tau(k+2, b) = k+2. \end{aligned}$$

This is shown in Fig. 1⁺. One verifies that \mathbf{A}_{k+1} is reduced.

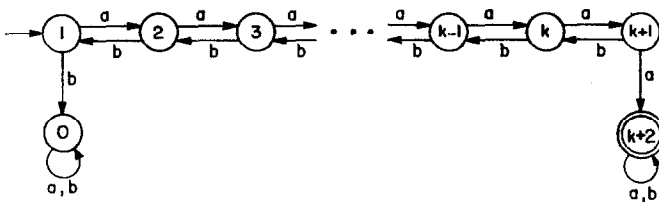


FIG. 1⁺. Automaton \mathbf{A}_{k+1} .

Before proceeding we will prove the following property of \mathbf{A}_{k+1} . Let

$$u_0 = (ab)^n,$$

and for $j \geq 1$ let

$$u_j = u_{j-1}(au_{j-1}bu_{j-1})^m \quad \text{and} \quad v_j = u_{j-1}(bu_{j-1}au_{j-1})^m,$$

be defined as in Lemma 3⁺, with $x = (ab)^n$, $y = a$ and $z = b$. Then

$$\begin{aligned} \tau(i, u_j) &= i & \text{for } 1 \leq i \leq k-j, \\ \tau(i, u_j) &= k+2 & \text{for } k-j+1 \leq i \leq k+1. \end{aligned} \quad (9)$$

We verify this claim by induction on j .

$j = 0$: This is easily verified for $u_0 = (ab)^n$.

$j > 0$: Assume that (9) holds for u_j . Denote by \underline{x} the transformation on the set Q of states of \mathbf{A}_{k+1} caused by x . The transformation $\underline{u_j}$ is as shown in the first row of Fig. 2⁺ by the inductive assumption. From Fig. 1⁺ it is easily verified that $\underline{u_j a}$, $\underline{u_j a u_j}$, and $\underline{u_j a u_j b}$ are as shown in Fig. 2⁺, and that

$$\underline{u_j a u_j b u_j} = \underline{u_j a u_j b} \quad (10)$$

and

$$\underline{u_j a u_j b u_j a} = \underline{u_j a u_j}.$$

Thus

$$\underline{u_j a u_j b u_j a (u_j b u_j)} = \underline{u_j a u_j u_j b u_j}.$$

Noting that $\underline{u_j u_j} = \underline{u_j}$, we have

$$\underline{u_j (a u_j b u_j)^2} = \underline{u_j (a u_j b u_j)}.$$

Hence

$$\underline{u_{j+1}} = \underline{u_j (a u_j b u_j)^m} = \underline{u_j (a u_j b u_j)}.$$

From (10) and Fig. 2⁺, we have the claim (9) for u_{j+1} .

	1	2	...	k-j-1	k-j	k-j+1	...	k	k+1
$\underline{u_j}$	1	2	...	k-j-1	k-j	k+2	...	k+2	k+2
$\underline{u_j a}$	2	3	...	k-j	k-j+1	k+2	...	k+2	k+2
$\underline{u_j a u_j}$	2	3	...	k-j	k+2	k+2	...	k+2	k+2
$\underline{u_j a u_j b}$	1	2	...	k-j-1	k+2	k+2	...	k+2	k+2

FIG. 2⁺. Transformations in \mathbf{A}_{k+1} .

PROPOSITION 5⁺. $L_{k+1} \in \mathcal{B}^{(k+1)} - \mathcal{B}^{(k)}$, i.e., L_{k+1} is a depth- $(k+1)$ language.

Proof. First we show that $L_{k+1} \notin \mathcal{B}^{(k)}$. By (9) $\tau(1, u_{k-1}) = 1$ and $\tau(2, u_{k-1}) = k+2$. Thus

$$\tau(1, u_k) = \tau(1, u_{k-1}(au_{k-1}bu_{k-1})^m) = k+2,$$

and

$$\tau(1, v_k) = 0.$$

Therefore $u_k \in L_{k+1}$ but $v_k \notin L_{k+1}$. By Lemma 3⁺ $u_k \sim_n^k v_k$. Hence L_{k+1} cannot be a union of congruence classes of \sim_n^k , and $L_{k+1} \notin \mathcal{B}^{(k)}$.

Next we will show that the language L_{k+1} recognized by \mathbf{A}_{k+1} is in $\mathcal{B}^{(k+1)}$. We will show in Lemma 4⁺ that a related language, D_k , is in $\mathcal{B}^{(k)}$. Let

$$\begin{aligned} D_0 &= 1, \\ D_k &= (aD_{k-1}b)^*, \quad \text{for } k \geq 1. \end{aligned}$$

One easily verifies that $D_k = \{w \in A^* \mid \tau(1, w) = 1\}$ in \mathbf{A}_{k+1} . Note also that

$$D_{k-1} \subseteq D_k \quad \text{for all } k \geq 1.$$

Let $Z_i = \{w \in A^* \mid \tau(1, w) = i\}$. Then:

$$\begin{aligned} Z_0 &= D_k b A^*, \\ Z_1 &= D_k, \\ Z_{i+1} &= Z_i a D_{k-i} \quad \text{for } 1 < i \leq k, \end{aligned}$$

and

$$L_{k+1} = Z_{k+2} = Z_{k+1} a A^* = (D_k a D_{k-1} a D_{k-2} a \cdots D_2 a D_1 a) a A^*, \quad (11)$$

for we have

$$\begin{aligned} Z_{k+1} &= Z_k a = Z_k a 1 = Z_k a D_0, \\ Z_k &= Z_{k-1} a (ab)^* = Z_{k-1} a D_1, \end{aligned}$$

etc. The claim that $L_{k+1} \in \mathcal{B}^{(k+1)}$ now follows from (11) if we assume Lemma 4⁺. ■

LEMMA 4⁺. For $k \geq 1$ let

$$\bar{E}_k = D_{k-1} b A^* \cup A^* b (b D_{k-1})^{k-1} b A^* \cup A^* a D_{k-1} \cup A^* a (D_{k-1} a)^{k-1} a A^*.$$

Then $E_k = D_k$, showing explicitly that $D_k \in \mathcal{B}^{(k)}$.

Proof. We verify:

- (a) $x \in D_{k-1} b A^*$ implies $\tau(1, x) = 0$.
- (b) $x \in A^* b$ implies $\tau(1, x) \neq k+1$. Hence $y \in (D_{k-1} b)^{k-1} b A^*$ implies $\tau(1, xy) \in \{0, k+2\}$.
- (c) $x \in A^* a D_{k-1}$ implies $\tau(1, x) \neq 1$.
- (d) $x \in A^* a (D_{k-1} a)^{k-1} a A^*$ implies $\tau(1, x) \in \{0, k+2\}$.

Therefore, we have shown that $x \in \bar{E}_k$ implies $x \in D_k$.

Conversely, if $x \in \bar{D}_k$ and $\tau(1, x) \in \{2, \dots, k+1\}$, then $x \in A^*aD_{k-1}$. Thus $x \in \bar{E}_k$. Next suppose $\tau(1, x) = 0$ and $x = x_1x_2$ implies $\tau(1, x_1) \neq k+1$. Then $x \in D_{k-1}bA^*$. Now suppose $\tau(1, x) = 0$ and x "goes through" $k+1$. Let x_1 be the longest prefix of x such that $\tau(1, x_1) = k+1$. Then x is of the form $x = x_1bx_2$ where $\tau(1, x_1b) = k$. Now $x_1b \in A^*b$ and

$$x_2 \in bD_1bD_2 \cdots bD_{k-1}bA^* \subseteq (bD_{k-1})^{k-1}bA^*.$$

Thus $x_1bx_2 \in A^*b(bD_{k-1})^{k-1}bA^*$ and $x \in \bar{E}_k$. Similarly we verify that $\tau(1, x) = k+2$ implies

$$x \in A^*a(D_{k-1}a)^{k-1}aA^*.$$

For let x_1 be the longest prefix of x such that $\tau(1, x_1) = 1$. Then x is of the form $x = x_1ax_2$, where

$$x_2 \in (D_{k-1}aD_{k-2}a \cdots D_1a)aA^* \subseteq (D_{k-1}a)^{k-1}aA^*.$$

Hence the claim holds and in all cases $x \in \bar{D}_k$ implies $x \in \bar{E}_k$. Therefore $\bar{D}_k \subseteq \bar{E}_k$ and the lemma follows. ■

This concludes the induction step and we can now state our main result:

THEOREM. *The dot-depth hierarchy of star-free languages is infinite.*

Proof. For each $k \geq 1$ we have exhibited a language L_{k+1} that is in $\mathcal{B}^{(k+1)} - \mathcal{B}^{(k)}$. ■

4+. ON SYNTACTIC SEMIGROUPS OF DEPTH- k LANGUAGES

We now generalize the notion of 1-mutativity. Let S be any semigroup and $k > 1$ an integer. S is *k-mutative* iff there exists $m \geq 1$ such that for each $f, g \in S$

$$h_{k-1}(fh_{k-1}gh_{k-1})^m = h_{k-1}(gh_{k-1}fh_{k-1})^m$$

where

$$h_1 = (fg)^m$$

and

$$h_k = h_{k-1}(fh_{k-1}gh_{k-1})^m \quad \text{for } k > 1.$$

The following is a necessary condition for membership in $\mathcal{B}^{(k)}$:

PROPOSITION 6+. *Let $L \subseteq A^+$ and let S_L be the syntactic semigroup of L .*

(a) *If $L \in \mathcal{B}^{(k)}$ then for each idempotent $e \in S_L$, $eS_L e$ is finite, aperiodic, and k -mutative.*

(b) *Suppose S_L is a monoid. Then $L \in \mathcal{B}^{(k)}$ implies S_L is finite, aperiodic, and k -mutative.*

Proof. (a) Suppose $L \in \mathcal{B}^k$. Then L is a union of congruence classes of \sim_n^k for some $n \geq 1$. Since \sim_n^k is of finite index, S_L is finite.

Let $f \in S_L$ and let $y \in A^+$ be such that $y = f$. By Lemma 1+

$$y^{m_k} \sim_n^k y^{m_k+1}.$$

Since L is a union of congruence classes of \sim_n^k it follows that

$$f^{m_k} = f^{m_k+1}. \quad (12)$$

Hence S_L is group free.

Now let e, f , and $g \in S_L$ be such that e is an idempotent and let $u, x, y, z \in A^+$ be such that $\underline{u} = e, \underline{y} = f, \underline{z} = g$, and $x = u^{n+1}$. By Lemma 3+

$$u_{k-1}(yu_{k-1}zu_{k-1})^{m_k} \sim_n^k u_{k-1}(zu_{k-1}yu_{k-1})^{m_k}.$$

Thus

$$\underline{u_{k-1}}(\underline{f}\underline{u_{k-1}}\underline{g}\underline{u_{k-1}})^{m_k} = \underline{u_{k-1}}(\underline{g}\underline{u_{k-1}}\underline{f}\underline{u_{k-1}})^{m_k}.$$

Now one easily verifies by induction on k that $\underline{u_k} = \underline{e}\underline{u_k}e$ for all $k \geq 0$. Thus

$$\underline{u_k} = \underline{u_{k-1}}((efe) \underline{u_{k-1}}(ege) \underline{u_{k-1}})^{m_k}.$$

Now let

$$h_1 = \underline{u_1} = e((efe) e(ege)e)^{m_k} = ((efe)(ege))^{m_k},$$

and

$$h_k = \underline{u_k} \quad \text{for } k > 1.$$

Then $\underline{u_k} = \underline{v_k}$ implies

$$h_{k-1}((efe) h_{k-1}(ege) h_{k-1})^{m_k} = h_{k-1}((ege) h_{k-1}(efe) h_{k-1})^{m_k}. \quad (13)$$

Now (a) follows from (12) and (13).

(b) Let 1 be the identity of S_L ; then (12) and (13) hold with $e = 1$. ■

Observe that the notion of k -mutativity defines an infinite hierarchy of finite semi-groups. This follows from the example in Fig. 1+, since the syntactic semigroup of A_{k+1} is $(k+1)$ -mutative, but not k -mutative.

ACKNOWLEDGMENT

The authors wish to thank K. Čulík II for suggesting the form of the expression in Lemma 4+.

REFERENCES

1. J. A. BRZOWSKI, Hierarchies of aperiodic languages, *Rev. Française d'Automatique Informat. Recherche Operationnelle Série Rouge (Informatique Théorique)* 10 (1976), 35-49.
2. J. A. BRZOWSKI, K. CULÍK II, AND A. GABRIELIAN, Classification of noncounting events, *J. Comput. System Sci.* 5 (1971), 41-53.
3. R. S. COHEN AND J. A. BRZOWSKI, Dot-depth of star-free events, *J. Comput. System Sci.* 5 (1971), 1-16.
4. R. MCNAUGHTON AND S. PAPERT, "Counter-Free Automata," M.I.T. Press, Cambridge, Mass., 1971.
5. M. P. SCHÜTZENBERGER, On finite monoids having only trivial subgroups, *Inform. Contr.* 8 (1965), 190-194.
6. I. SIMON, Piecewise testable events, in "Lecture Notes in Computer Science 33," Springer-Verlag, New York, 1975.